Statistical Research Report
Institute of Mathematics
University of Oslo

No. 3
July 1965

ON ASYMPTOTICALLY NORMAL ESTIMATORS

by

Grete Usterud Fenstad

1. Introduction

Let X_1, X_2, ... be independent and identically distributed observations with probability densities $p(x;\theta)$ with respect to some measure μ_0. Extending some work of Barankin and Gurland [2] we shall in this paper first give a proof of the existence of consistent, asymptotically normal distributed (c.a.n.) estimators under relative mild restrictions on the family $\mathcal{P} = \{p(x;\theta)\}$ of probability densities. Instead of assuming, as did Barankin and Gurland, for the existence proof that so-called separators exist, we shall be able to rely on some results of differential geometry, viz. the existence of tubular neighborhoods, in carrying out the proof. In actually constructing c.a.n. estimators separators are of great use, but they are irrelevant for the general existence theorem.

Chiang [3] has done some work in the same direction, but he puts the restrictions on the estimators instead of on the family $\mathcal{P}$ of probability densities.

The asymptotic covariance matrix depends on two variables, with respect to which we want to minimize the matrix. In section 3 we shall obtain within our setting a lower bound for this matrix keeping one of the variables fixed, and it will be shown that in a special case this lower bound always is obtained. Our work here is in addition to [2] also closely related to Chiang [3] and we shall be very brief in our exposition.

Imposing some further restrictions on the family $\mathcal{P}$ we shall in the last section of this paper obtain an absolutely lower bound for the asymptotic covariance matrix. This section generalizes somewhat the results of the corresponding part of the work of Barankin and Gurland.

In a certain sense this paper is an exposition of the basic ideas in Barankin and Gurland [2]. We believe, however, that the technical simplifications made in this paper makes the theory easier accessible and thus should be of interest.

2. Existence of c.a.n. estimators

Let X be a random vector, taking values in the m-dimensional Euclidean space $\mathbb{R}^m$, with probability density $p(x;\theta)$ with respect to a measure μ_o over $\mathbb{R}^m$. The parameter θ is supposed to belong to an open k-dimensional set $\Theta \subset \mathbb{R}^k$.

We shall consider as basic for our study the following class of families of probability densities:

DEFINITION 1. The family $\mathcal{P}$ of probability densities belongs to the class Π if

(i) there exists a finite set of $s \geqslant k$ μ_o-measurable, real-valued functions on $\mathbb{R}^m$, $\Phi = (\Phi_1, \ldots, \Phi_s)$, such that $E\Phi_i(X) = A_i(\theta)$ and $\operatorname{cov}(\Phi_i(X), \Phi_j(X)) = a_{ij}(\theta)$, $i,j = 1, \ldots, s$ exist and are finite for every $\theta \in \Theta$.

(ii) the rank of the matrix

$$\dot{A}(\theta) = \begin{pmatrix} \frac{\partial A_1(\theta)}{\partial \theta_1} & \frac{\partial A_2(\theta)}{\partial \theta_1} & \cdots & \frac{\partial A_s(\theta)}{\partial \theta_1} \\ - & - & - & - \\ \frac{\partial A_1(\theta)}{\partial \theta_k} & \frac{\partial A_2(\theta)}{\partial \theta_k} & \cdots & \frac{\partial A_s(\theta)}{\partial \theta_k} \end{pmatrix}$$

is k for every $\theta \in \Theta$

(iii) the mapping A of Θ into $\mathbb{R}^s$ defnined by

$$A(\theta) = (A_1(\theta), \ldots, A_s(\theta))$$

is homeomorphic and continuously differentiable on Θ.

(iv) <u>there exists a symmetric, positive definite, s x s matrix</u>

$$d(\theta) = (d_{ij}(\theta))$$

<u>such that</u>

$$\dot{A}(\theta)\ d(\theta)\ a(\theta) = \dot{A}(\theta)$$

<u>for every $\theta \in \Theta$, and where $a(\theta) = (a_{ij}(\theta))$</u>

Condition (iv) in the definition generalizes the usual condition that $a(\theta)$ is non-singular. If $a(\theta)$ is non-singular one may choose $d(\theta) = a(\theta)^{-1}$. The idea of this extension has been found in a paper of Ferguson [6].

Let $X_1, X_2, \ldots$ be independent observations of X, and set $Y_n = (X_1, \ldots, X_n)$. The random variable Y_n has probability density $\prod_{i=1}^{n} p(x_i, \theta)$ with respect to the product measure $\mu_n = \mu_o \times \ldots \times \mu_o$.

DEFINITION 2. <u>An estimator for θ, F_n, is a μ_n-measurable function of Y_n to Θ.</u>

We are going to prove the existence of an estimator for θ which depends on Y_n only through

$$B_n(Y_n) = \left(\frac{1}{n}\sum_{i=1}^{n}\phi_1(X_i), \ldots, \frac{1}{n}\sum_{i=1}^{n}\phi_s(X_i)\right)$$

We shall in the following denote the image of Θ under A by R and elements in $\mathbb{R}^s$ by z. Further, a mapping is an element of C' if it is continuously differentiable. A neighborhood is always supposed to be an open set.

LEMMA 1. <u>If the family of probability densities $\mathcal{P}$ belongs to the class Π, there exists a neighborhood S of R in $\mathbb{R}^s$ and a mapping H, $H \in C'$, of S to Θ, such that $A(H(z)) = z$ for every $z \in \mathbb{R}^s$</u>.

<u>Proof</u>. The proof is an immediate application of the theorem on the existence of tubular neighborhoods (see e.g. Munkres [7, 5.5 Theorem]) which essentially states:

Let f be a homeomorphic, continuously differentiable mapping from an open set M to $\mathbb{R}^m$. The Jacobian of f has maximal rank in every point of M. Then there is a neighborhood W of $f(M)$ and a continuously differentiable mapping r of W to $f(M)$ such that $r(y) = y$ for every $y \in f(M)$.

Our mapping A of Θ to $\mathbb{R}^s$ satisfies the conditions of the theorem. We therefore have the existence of a neighborhood S of $R = A(\Theta)$ and a mapping r, $r \in C'$, of S to R such that $r(z) = z$ for every $z \in R$.

To prove the lemma, define $H = A^{-1}r$. H is a mapping of S to Θ, and $A(H(z)) = r(z) = z$ for every $z \in R$. To complete the proof we have to show that $H \in C'$. For this purpose we apply the rank theorem [5; 10.3.1]. Let $z \in S$, and $\theta \in \Theta$ such that $A(\theta) = r(z)$. Then there exists a neighborhood U of θ and a neighborhood $V \supset A(U)$ of $r(z)$ such that $A = v A_o u$, where u is a homeomorphic mapping of U to the open k-dimensional unit ball $I^k = \{(\theta_1, \ldots, \theta_k); |\theta_i| < 1, i = 1, \ldots, k\}$ and v is a homeomorphic mapping of I^s to V. u and v and their inverse mappings are elements of C'. A_o is the mapping of I^k to I^s defined by $A_o(\theta_1, \ldots, \theta_k) = (\theta_1, \ldots, \theta_k, 0, \ldots, 0)$. Because of the continuity of r there exists a neighborhood V_1 of z, such that $r(V_1) \subset V \cap R$. Now we can write $H = A^{-1}r$, i.e.

$$H = u^{-1} A_o^{-1} v^{-1} r,$$

on the neighborhood V_1, where A_o^{-1} is the mapping of I^s to I^k defined by $A_o^{-1}(\theta_1, \ldots, \theta_k, \ldots, \theta_s) = (\theta_1, \ldots, \theta_k)$. Obviously $A_o^{-1} \in C'$. Thus $H \in C'$ on a neighborhood V_1 of z because u^{-1}, A_o^{-1}, v^{-1}, $r \in C'$ on their respective domains. Since $z \in S$ is chosen arbitrarily the proof is completed. //

Denote by <u>$\mathcal{H}(\mathcal{P}, \Phi)$</u> the collection of all mappings $H \in C'$ defined on some neighborhood of R with values in Θ which can be written in the form $H = A^{-1}r$, where $r \in C'$ is a mapping of the domain of H to R satisfying $r(z) = z$ for every $z \in R$. Lemma 1 tells us that $\mathcal{H}(\mathcal{P}, \Phi)$ is non-empty.

THEOREM 1. <u>For every $\mathcal{P} \in \Pi$ there exists a c.a.n. estimator for θ.</u>

<u>Remark</u>. This corresponds to Theorem 4.2 in [2]. The proof is quite similar, but for completeness we sketch the argument without going into details.

<u>Proof</u>. Since $\mathcal{P} \in \Pi$, there exists a neighborhood S of R and a $H \in \mathcal{H}(\mathcal{P}, \Phi)$. Set

$$F_n(Y_n) = \begin{cases} H(B_n(Y_n) & \text{if } B_n(Y_n) \in S \\ \theta^o & \text{if } B_n(Y_n) \notin S \end{cases}$$

where θ^o is an arbitrary element of Θ. One easily verifies that F_n is an estimator.

We shall prove that $F_n(Y_n)$ is asymptotically normal distributed with mean θ. From this follows that $F_n(Y_n)$ is consistent.

From the well known asymptotic distribution of $\left[B_n(Y_n)-A(\theta)\right]$ it follows that $\left[B_n(Y_n) - A(\theta)\right] \dot{H}(\theta)$, where

$$\dot{H}(\theta) = \left(\frac{\partial H(z)}{\partial z}\right)_{z=A(\theta)} = \begin{pmatrix} \frac{\partial H_1(z)}{\partial z_1} & \frac{\partial H_2(z)}{\partial z_1} & \cdots & \frac{\partial H_k(z)}{\partial z_1} \\ - - - & - - - & - - - & - - - \\ \frac{\partial H_1(z)}{\partial z_s} & \frac{\partial H_2(z)}{\partial z_s} & \cdots & \frac{\partial H_k(z)}{\partial z_s} \end{pmatrix}_{z=A(\theta)},$$

is asymptotically normal distributed with mean 0 and covariance matrix $\frac{1}{n} \dot{H}(\theta)' a(\theta)\dot{H}(\theta)$.

Since $B_n(Y_n)$ converges to $A(\theta) \in S$ in probability it follows that $F_n(Y_n) - H(B_n(Y_n)) = \left[F_n(Y_n) - \theta\right] - \left[H(B_n(Y_n)) - \theta\right]$ converges to 0 in probability. As $H \in C'$ we may write

$$H(B_n(Y_n)) - \theta = \left[B_n(Y_n) - A(\theta)\right]\left(\frac{\partial H(z)}{\partial z}\right)_{z=z'}$$

where z' lies in the interior of the line segment joining $B_n(Y_n)$ and $A(\theta)$. As n tends to infinity, z' converges to $A(\theta)$ in probability, thus, since $H \in C'$, $\left[H(B_n(Y_n)) - \theta\right] - \left[B_n(Y_n) - A(\theta)\right] \dot{H}(\theta)$ converges to 0 in probability.

It has now been verified that $\left[F_n(Y_n) - \theta\right] - \left[B_n(Y_n)-A(\theta)\right]\dot{H}(\theta)$ converges to 0 in probability, and it follows that $F_n(Y_n) - \theta$ and $\left[B_n(Y_n) - A(\theta)\right] \dot{H}(\theta)$ have the same asymptotic distributions, since we have already shown that $\left[B_n(Y_n) - A(\theta)\right] \dot{H}(\theta)$ has a limiting distribution [2; Lemma 4.1]. Their common asymptotic distribution is the normal distribution with mean 0 and covariance matrix $\frac{1}{n}\dot{H}(\theta)'a(\theta)\dot{H}(\theta)$, thus $F_n(Y_n)$ is asymptotically normal distributed with mean θ and covariance matrix $\frac{1}{n}\dot{H}(\theta)'a(\theta)\dot{H}(\theta)$. //

3. Lower bound of the asymptotic covariance matrix when Φ is considered fixed.

Let $H \in \mathcal{H}(\mathcal{P}, \Phi)$, we shall say that an estimator $F_n(Y_n)$ is generated by the method $\mathcal{M}$ (or $\mathcal{M}_H$ if we want to indicate the particular $H \in \mathcal{H}(\mathcal{P}, \Phi)$ which is used) if the estimator satisfies the equality $F_n(Y_n) = H(B_n(Y_n))$ for points $B_n(Y_n) \in S$, where S is the domain of H. The value of $F_n(Y_n)$ when $B_n(Y_n) \notin S$ is of no consequence.

We have just proved that $\mathcal{M}_H$ generates c.a.n. estimators with asymptotic covariance matrix $\frac{1}{n}\dot{H}(\theta)'\, a(\theta)\, \dot{H}(\theta)$. But the covariance matrix depends on the choice of Φ via both $\dot{H}(\theta)$ and $a(\theta)$, and on the choice of H. We want to minimize the matrix in a certain sense over both H and Φ .

DEFINITION 3. The matrix A is greater than or equal to B $(A \geqslant B)$ if $A - B$ is positive semi-definite.

If A and B are two covariance matrices of same order and $A \geqslant B$, it follows that the variances of A is greater than or equal to the corresponding variances of B.

THEOREM 2. If $\mathcal{P} \in \Pi$, then the asymptotic covariance matrix of an estimator generated by $\mathcal{M}$ is always greater than or equal to

$$\frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1}$$

for every $\theta \in \Theta$.

Proof. The θ will be surpressed to simplify the notation. Since d is assumed to be positive definite and the rank of $\dot{A}$ is k, $\dot{A}\, d\, \dot{A}'$ is positive definite, and its inverse matrix exists.

a is a covariance matrix, therefore positive semi-definite, hence

$$v\left[(\dot{H}' - (\dot{A}d\dot{A}')^{-1}\dot{A}d)a(\dot{H}-d\dot{A}'(\dot{A}d\dot{A}')^{-1}\right]v'$$
$$= \left[v(\dot{H}' - (\dot{A}d\dot{A}')^{-1}\dot{A}d)\right] a \left[v(\dot{H}' - (\dot{A}d\dot{A}')^{-1}\dot{A}d)\right]' \geqslant 0$$

where $v = (v_1, \ldots, v_k)$ is an arbitrary k-dimensional vector. According to Definition 3 we thus have

$$(\dot{H}' - (\dot{A}d\dot{A}')^{-1}\dot{A}d)\; a\; (\dot{H} - d\dot{A}'(\dot{A}d\dot{A}')^{-1}) \geqslant 0$$

This gives $\dot{H}' a \dot{H} \geqslant (\dot{A}d\dot{A}')^{-1}$, when we substitute $\dot{A}$ for $\dot{A}da$ and I for $\dot{H}'\dot{A}'$ in the above expression. (The equality $\dot{H}'\dot{A}' = I$ follows by differentiating $H(A(\theta)) \equiv \theta$.) //

The method of proof is similar to proofs given in Ferguson [6] and Chiang [3].

The question if there exists a $H \in \mathcal{H}(\mathcal{P}, \Phi)$ such that the asymptotic covariance matrix $\frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1}$ is obtained by the method $\mathcal{M}_H$ still remains to be answered. A partial answer is the following theorem.

THEOREM 3. <u>If $\mathcal{P} \in \Pi$ and $s = k$, every estimator generated by $\mathcal{M}$ will have asymptotic covariance matrix</u>.

$$\frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1}$$

<u>Proof</u>. In the case $s = k$, one may verify by use of the inverse function theorem that R itself is open. We therefore may choose $S = R$ and $H = A^{-1}$ on R. Since $A(H(z)) = z$ for every $z \in R$ we obtain by differentiating that $\dot{A}(\theta)'\,\dot{H}(\theta)' = I$ or $\dot{H}(\theta)' = (\dot{A}(\theta)')^{-1}$. The condition $\dot{A}(\theta)d(\theta)a(\theta) = \dot{A}(\theta)$ gives $a(\theta) = d(\theta)^{-1}$ so $a(\theta)$ must be non-singular. Thus, the covariance matrix of an estimator generated by $\mathcal{M}_H$ is

$$\frac{1}{n}\,\dot{H}(\theta)'a(\theta)\dot{H}(\theta) = \frac{1}{n}(\dot{A}(\theta)')^{-1}d(\theta)^{-1}\dot{A}(\theta)^{-1} = \frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1} \quad //$$

Methods for constructing mappings $H \in \mathcal{H}(\mathcal{P}, \Phi)$ such that $\mathcal{M}_H$ generates estimators for θ with asymptotic covariance matrix $\frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1}$ have been considered in several papers; for example by minimizing quadratic forms with respect to θ in [8], [2], [9], [3]; by minimizing quadratic forms under linearized restrictions in [8], [2], [3], or by solving a linear form with respect to θ in [6].

4. The minimal asymptotic covariance matrix.

In order to obtain a lower bound for the asymptotic covariance matrix we need more detailed information about the matrix $d(\theta)$ (see Definition 1 (iv)). $d(\theta)$ is not uniquely determined by the condition $\dot{A}(\theta)d(\theta)a(\theta) = \dot{A}(\theta)$, however, two matrices $d_1(\theta)$ and $d_2(\theta)$ both satisfying the condition will give the same value of the expression $\dot{A}(\theta)d(\theta)\dot{A}(\theta)'$ as the following calculation shows:

$$\dot{A}(\theta)d_1(\theta)\dot{A}(\theta)' = \dot{A}(\theta)d_2(\theta)a(\theta)d_1(\theta)\dot{A}(\theta)' = \dot{A}(\theta)d_2(\theta)\dot{A}(\theta)' .$$

Another useful property is contained in the following lemma.

LEMMA 2. <u>$d(\theta)$ may always be chosen such that $d(\theta)a(\theta)$ has only 0 and 1 as eigenvalues</u>.

<u>Proof</u>. We assume the existence of a d satisfying $\dot{A}da = \dot{A}$, and that d is symmetric, positive definite. Thus there exists a non-singular matrix c, such that $d = c'c$. Since $|da - \lambda I| = |cac' - \lambda I|$, we may instead prove that c can be chosen such that cac' has only 0 and 1 as eigenvalues.

As cac' is symmetric there exists an orthogonal matrix P such that

$$Pcac'P' = D$$

where D is a diagonal matrix with the eigenvalues D_i, $i = 1, \ldots, s$ of cac' on the diagonal. We define a new diagonal matrix Q in the following way

$$Q_i = \begin{cases} 1 & \text{if } D_i = 0 \\ \frac{1}{\sqrt{D_i}} & \text{if } D_i \neq 0 \end{cases} \qquad i = 1, \ldots, s$$

where Q_i, $i = 1, \ldots, s$ are the diagonal elements.

Let the rank of a be equal to t, then $(s-t)$ of the eigenvalues of cac' is equal to 0.

Further observe that $QDQ' = I_t$ where I_t is a diagonal matrix such that t of the elements on the diagonal equal 1, the rest 0, and that $D \cdot I_t = D$.

We shall prove that

$$c_o a c_o' \quad ,$$

where

$$c_o = QPc \quad ,$$

has only 0 and 1 as eigenvalues, and that

$$d_o = c_o' c_o$$

satisfies the equation

$$\dot{A} da = \dot{A} \ .$$

Since the characteristic polynomial of $c_o a c_o'$ may be written as

$$|c_o a c_o' - \lambda I| = |QPcac'P'Q' - \lambda I| = |I_t - \lambda I|$$

the first assertion is proved

From

$$\dot{A} da = \dot{A} c'ca = \dot{A}$$

it follows that

$$\dot{A} c'P'Q^{-1} I_t = \dot{A} c'P'Q'$$

by inserting $P'Q^{-1}QP = I$ and multiplying on the right by $c'P'Q'$, or

(4.1) $$\dot{A} c_o' D = \dot{A} c_o'$$

noticing that $D = Q^{-1} I_t Q^{-1}$ and that diagonal matrices commute. We now do the following calculations using (4.1) twice:

$$\dot{A} d_o a c_o' = \dot{A} c_o' c_o a c_o' = \dot{A} c_o' I_t$$

$$= \dot{A} c_o' D \cdot I_t = \dot{A} c_o' D = \dot{A} c_o'$$

By multiplying on the right by $c_o'^{-1}$ we in the end obtain

$$\dot{A} d_o a = \dot{A}$$

which was to be proved. //

The lower bound $\frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1}$ obtained in Theorem 2 depends on the choice of Φ. Imposing further regularity assumptions on the class Π, we propose to show that this bound is always greater than or equal to a matrix independent of Φ.

DEFINITION 4. <u>A family of probability densities $\mathcal{P}$ belongs to Π_o, if $\mathcal{P} \in \Pi$ and</u>

(i) <u>we may in</u>

$$\int p(\cdot\,;\theta)d\mu_o$$

<u>and</u>

$$\int \Phi_i(\cdot)p(\cdot;\theta)d\mu_o \qquad i = 1, \ldots, s$$

<u>differentiate with respect to $\theta_\varkappa$, $\varkappa = 1, \ldots, k$ under the integral signs</u>

(ii) <u>the integrals</u>

$$\int \left(\frac{\partial \ln p(\cdot;\theta)}{\partial \theta_\varkappa}\right)^2 p(\cdot;\theta)d\mu_o \qquad \varkappa = 1, \ldots, k$$

<u>are finite for every $\theta \in \Theta$</u>

(iii) <u>the matrix</u>

$$\rho(\theta) = (\rho_{\varkappa\lambda}(\theta))_{\varkappa,\lambda = 1, \ldots, k},$$

<u>where</u> $$\rho_{\varkappa\lambda}(\theta) = \int \frac{\partial \ln p(\cdot;\theta)}{\partial \theta_\varkappa} \frac{\partial \ln p(\cdot;\theta)}{\partial \theta_\lambda} p(\cdot;\theta)d\mu_o,$$

<u>is non-singular for every $\theta \in \Theta$</u>.

These conditions are similar to those of Barankin and Gurland [2], but they restrict themselves to the class Π_o all through their paper. Here Theorems 1, 2, 3 are valid in the larger class Π.

The main point in the proof of the next theorem is [2; Lemma 2.1] which essentially says:

<u>If $\mathcal{P} \in \Pi_0$ and if $\Phi = (\Phi_1, \ldots, \Phi_s)$ is a set of functions satisfying the conditions of Definition 1, then</u>

$$p(x;\theta) = \exp(\alpha_0(\theta) + \Phi_0(x) + \sum_{i=1}^{s} \alpha_i(\theta)\Phi_i(x)) \tag{4.2}$$

<u>for every $\theta \in \Theta$, is necessary and sufficient condition that</u> $\frac{\partial \ln p(x;\theta)}{\partial \theta_{\kappa}}$ <u>$\kappa = 1, \ldots, k$ for every $\theta \in \Theta$, can be written in the form</u>

$$\frac{\partial \ln p(x;\theta)}{\partial \theta_{\kappa}} = \sum_{i=1}^{s} \gamma_{\kappa i}(\theta) \; (\Phi_i(x) - A_i(\theta)) \; . \tag{4.3}$$

Let F_n be any estimator of θ generated by the method $\mathcal{M}$ We are going to show that a necessary condition that F_n has a minimal asymptotic covariance matrix is that $\frac{\partial \ln p(x;\theta)}{\partial \theta_{\kappa}}$ $\kappa = 1, \ldots, k$ is of the form (4.3) or equivalent, that $p(x;\theta)$ is of the form (4.2).

Define as usual the inner product of h and k to be $\langle h,k \rangle = \int h(.)k(.)p(.;\theta)d\mu_0$ where h and k are square-integrable, real-valued functions on $\mathbb{R}^m$

Let $d = c'c$ be a version of d having the properties (i) $\dot{A}d\dot{a} = \dot{A}$ and (ii) $cac' = I_t$. Define $\psi = (\psi_1, \ldots, \psi_s)$ by

$$\psi(x;\theta)' = c(\theta) \; (\Phi(x) - A(\theta))'$$

Since

$$\int \psi(.;\theta)' \; \psi(.;\theta)p(.;\theta)d\mu_0 = c(\theta)a(\theta)c(\theta)' = I_t \; , \qquad \psi_1, \ldots, \psi_t$$

form an orthonormal set in L_2.

We now proceed as in [2; pp.115-116], the calculations are carried out in detail there. We may write

$$\frac{\partial \ln p}{\partial \theta} = \bar{p}_{\kappa} + \hat{p}_{\kappa} \qquad \kappa = 1, \ldots, k$$

where $\bar{p}_{\kappa}$ lies in the space spanned by $\psi_1, \ldots, \psi_t$ and $\hat{p}_{\kappa}$ in the orthogonal complement of this space. Since $\psi_1, \ldots, \psi_t$ form an orthonormal set, we obtain

$$(4.4)\quad \sum_{i=1}^{t} \left\langle \frac{\partial \ln p}{\partial \theta_{\varkappa}}, \psi_i \right\rangle \left\langle \frac{\partial \ln p}{\partial \theta_{\lambda}}, \psi_i \right\rangle =$$

$$= \left\langle \frac{\partial \ln p}{\partial \theta_{\varkappa}}, \frac{\partial \ln p}{\partial \theta_{\lambda}} \right\rangle - \left\langle \hat{p}_{\varkappa}, \hat{p}_{\lambda} \right\rangle = \rho_{\varkappa\lambda} - \left\langle \hat{p}_{\varkappa}, \hat{p}_{\lambda} \right\rangle$$

$$\varkappa, \lambda = 1, \ldots, k$$

We are interested in the matrix

$$\dot{A}d\dot{A}' = \dot{A}dad\dot{A}' = \dot{A}c' \, I_t(\dot{A}c')' \, .$$

Some calculations (see [2]) and use of (4.4) give us

$$(\dot{A}d\dot{A}')_{\varkappa\lambda} = \sum_{i=1}^{t} (\dot{A}c')_{\varkappa i} (\dot{A}c')_{\lambda i} = \sum_{i=1}^{t} \left\langle \frac{\partial \ln p}{\partial \theta_{\varkappa}}, \psi_i \right\rangle \cdot \left\langle \frac{\partial \ln p}{\partial \theta_{\lambda}}, \psi_i \right\rangle$$

$$= \rho_{\varkappa\lambda} - \left\langle \hat{p}_{\varkappa}, \hat{p}_{\lambda} \right\rangle \qquad \varkappa, \lambda = 1, \ldots, k$$

Let $v = (v_1, \ldots, v_k)$ be an arbitrary vector in $\mathbb{R}^k$, then

$$v \, \dot{A}d\dot{A}' \, v' \doteq v\rho v - \sum_{\varkappa,\lambda=1}^{k} v_{\varkappa} v_{\lambda} \left\langle \hat{p}_{\varkappa}, \hat{p}_{\lambda} \right\rangle$$

$$= v\rho v' - \left\langle \sum_{\varkappa=1}^{k} v_{\varkappa} \hat{p}_{\varkappa}, \sum_{\lambda=1}^{k} v_{\lambda} \hat{p}_{\lambda} \right\rangle \leqslant v\rho v'$$

Using Definition 3 this means that $\rho \geqslant \dot{A}d\dot{A}'$, and we may further conclude that

$$(\dot{A}d\dot{A}')^{-1} \geqslant \rho^{-1} \, ,$$

using [2; Lemma 6.5]. Equality obtains if and only if $\sum_{\varkappa=1}^{k} v_{\varkappa} \hat{p}_{\varkappa} = 0$ for every v. This implies that $\frac{\partial \ln p}{\partial \theta_{\varkappa}}$ $\varkappa = 1, \ldots, k$ must lie in the space spanned by $\psi_1, \ldots, \psi_t$, i.e. there exist $b_{\varkappa}(\theta) =$
$(b_{\varkappa 1}(\theta), \ldots, b_{\varkappa t}(\theta))$ $\varkappa = 1, \ldots, k$, or $b^{o}(\theta) =$
$(b_{\varkappa 1}(\theta), \ldots, b_{\varkappa t}(\theta), 0, \ldots, 0)$ $\varkappa = 1, \ldots, k$ such that

$$\frac{\partial \ln p(.;(\theta)}{\partial \theta_{\varkappa}} = b^{o}_{\varkappa}(\theta)\,\psi(.;\theta)' = b^{o}_{\varkappa}(\theta)c(\theta)(\phi(.) - A(\theta))' \quad \varkappa = 1,\ldots,k$$

This is of the form (4.3)

Estimators with asymptotic covariance matrix $\frac{1}{n}\rho(\theta)^{-1}$ for every $\theta \in (H)$ according to established terminology is said to be <u>uniformly efficient</u>.

Collecting our results we may now state the following theorem:

THEOREM 4. <u>Let $\mathcal{P}$ belong to the class Π_o, then</u>

$$\frac{1}{n}\dot{H}(\theta)'a(\theta)\dot{H}(\theta) \geqslant \frac{1}{n}\left[\dot{A}(\theta)d(\theta)\dot{A}(\theta)'\right]^{-1} \geqslant \frac{1}{n}\rho(\theta)^{-1}$$

<u>for every $\theta \in (H)$.</u>

<u>A necessary and sufficient condition for the method $\mathcal{M}_H$ to generate uniformly efficient estimators is that $\dot{H}(\theta)'a(\theta)\dot{H}(\theta) = (\dot{A}(\theta)d(\theta)\dot{A}(\theta)')^{-1}$ and that $p \in \mathcal{P}$ is of the form</u>

$$p(x;\theta) = \exp(\alpha_o(\theta) + \phi_o(x) + \sum_{i=1}^{s} \alpha_i(\theta)\phi_i(x))$$

<u>for every $\theta \in (H)$.</u>

That the condition of the theorem also is sufficient is immediate form the remarks above.

The method $\mathcal{M}$ does not generate uniformly efficient estimators when the probability densities are not of the form (4.2). Since there exist such estimators also in other cases, the method is not universal. For example, the sample median is a c.a.n. estimator and uniformly efficient if $p(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$ [4]. The limitation of the method $\mathcal{M}$ is that the estimators are supposed to be functions of the averages, $B_n(Y_n)$ (see remark immediately following Definition 2).

REFERENCES

[1] T.M. Apostol (1960): "Mathematical Analysis", Addison-Wesley Publishing Company

[2] E.W. Barankin and J. Gurland (1951): "On asymptotically normal, efficient estimators: I" University of California Publications in Statistics, Vol.1, No.6, pp.89-130

[3] C.L. Chiang (1956): "On regular best asymptotically normal estimates", Ann.Math.Stat., Vol.57, pp.336-351

[4] H.E. Daniels (1961): "The asymptotic efficiency of a maximum likelihood estimator", Fourth Berkeley Symposium, Vol.1, pp.151-163

[5] J. Dieudonné (1960): "Foundation of Modern Analysis", Academic Pre

[6] T.S. Ferguson (1953): "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities", Ann.Math.Stat., Vol.29, pp.1046-1062

[7] J.R. Munkres (1963): "Elementary differential topology", Annals of Mathematical Studies, Number 54, Princeton University Press

[8] J. Neyman (1949): "Contribution to the theory of the χ^2-test", Berkeley Symposium on Mathematical Statistics and Probability pp.239-273

[9] W.F. Taylor (1953): "Distance functions and regular best asymptotically normal estimates", Ann.Math.Stat., Vol.24, pp.85-92

**** ****
*****